

## **Problem Identification**

- **Problem Statement Formation:**

A NFL team needs help in assessing the Quarterback position with more information than the usual Passer Rating. The QB position is arguably the most important position on the field as they touch the ball on nearly every play. Their decisions can have a huge effect on the outcome of the game. The team has to evaluate the QBs on whom to offer a contract, bench or start.

- **Context:**

The team needs to make a decision on which QB to draft or who is the best one on the team. Also to predict their future QB rating based upon historical data. They also need to determine which factors have the largest impact on evaluating QBs.

- **Criteria for Success:**

Finding which elements that have a high correlation with success at the QB position. These factors will help train and find the best QBs in the draft or free agency.

- **Scope of Solution Space:**

The information from our project will be useful in predicting who can potentially be a good QB based on historical data. The 15 variables from our data will determine how to evaluate Quarterbacks.

- **Constraints:**

Not accounting for a variable that is unknown to us that plays a big role in evaluating the position. The final data might be too closely related to the old QB rating and it might not be even more helpful to the decision makers.

- **Stakeholders:**

Owner of team, President, General Manager, Director of Player Personnel, Director of Pro Personnel, Director of Pro Scouts, Director of College Scouts, Director of Administration and Head Scout.

- **Data Sources:**

Attempts, completions, yards, yards per attempt, touchdowns, interceptions, longest throw, how many points the team scored, sack and loss of yards. I will also add a few variables: completion percentage, touchdown to interception ratio,

yards per completion, longest throw per completion and longest throw per attempt. This is a 1.51mb data file from Kaggle.

I will create this analysis by wrangling and cleaning the data first. Then I will use graphs, scatter plots, heat maps, etc. to find the correlation between the factors and what makes a quarterback good or bad. I will formulate a predictive model in python. Supervised and unsupervised machine learning, like a random forest model or cluster analysis, is how I will predict which factors have the greatest impact on a quarterback's performance. The deliverables will be code to clean and analyze the data. I will deliver the new rating via a slidedeck, project report in github.